BEBR
FACULTY WORKING
PAPER NO. 90-1647

Predicting Appliance Ownership Using Logit, Neural Network, and Regression Tree Models

Chung-Ming Kuan
Halbert White

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 90-1647

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

April 1990

Predicting Appliance Ownership Using Logit,
Neural Network, and Regression Tree Models

Chung-Ming Kuan
Department of Economics
University of Illinois, Champaign, IL

and

Halbert White
University of California, San Diego

The authors are grateful for support by the California Energy Commission and the University Energy Research Group of the University of California.

ABSTRACT

In this paper we study forecasting performance of the logit model, a feedforward neural network model, and the regression tree model. These models are applied to predict household appliance stocks using the Miracle data sets collected by San Diego Gas and Electricity. Both in-sample and out-of-sample forecasting performance of each of these models are investigated. We find that the neural network model and the regression tree model exhibit clear advantages relative to the standard logit approach.

1. Introduction

Appliance saturation plays an important role in determining residential energy demand. In the short run, energy consumption of a household is a function of relevant socioeconomic and demographic variables, conditional on the appliance portfolio owned by that household. This motivates economists to study energy consumption using conditional demand functions, e.g., Parti and Parti (1980). In the long run, the consumer may be willing to pay a higher capital cost (in terms of discounted purchase price) for more efficient appliances in order to reduce the operating cost (in terms of energy price). This leads to the approach that models the demand for energy and choice of appliances simultaneously, e.g., Hausman (1979) and Dubin and McFadden (1984). However, this approach can handle only a small number of appliances. A simultaneous model of the demand for energy and the demand for a general appliance portfolio is usually intractable empirically.

In this paper we confine ourselves to the short run and estimate household appliance ownership probabilities conditional on household characteristics. A successful prediction of household appliance stocks should be helpful in improving short–run forecasts of residential energy demand. Here we apply three different models to the Miracle data sets collected by San Diego Gas and Electricity (SDG&E). These models are: the logit model, a feedforward neural network model (Rumelhart, Hinton, and Williams (1986)), and the regression tree model (Breiman, Friedman, Olshen, and Stone (1984)). The logit model is a typical approach in econometrics to dealing with discrete choice problems. The other two approaches are novel in the present context. Neural network modelling techniques have been widely used in the sciences recently and are known to be useful in performing complicated pattern recognition and classification tasks (e.g., Lapedes and Farber (1987a,b)). The regression tree analysis is a nonparametric statistical method specifically designed for classification problems. Our results investigate both the in–sample an out–of–sample forecasting performance of each of these models. The two novel approache exhibit clear advantages relative to the standard logit approach.

This paper proceeds as follows. In section 2, we discuss the methodologies use for estimating appliance ownership models. In section 3, we describe the dat characteristics and computer programs used for estimation. In section 4, we compare th performance of the three models. Section 5 concludes the paper.

2. Methodologies

Let $\{y_i\}$ be a sequence of independently distributed appliance ownershi dummy variables, where $y_i = 1$ if an appliance is owned by household i and $y_i =$ otherwise, and let X_i be a (column) vector of demographic variables (including a constar term) for household i. We are interested in forecasting appliance ownership conditional o X_i. We write

$$y_i = E[y_i|X_i] + \epsilon_i \,, \tag{1}$$

where $E[y_i|X_i]$ is the expectation of y_i conditional on X_i. It is clear that $E[y_i|X_i] =$ $P\{y_i=1|X_i\}$, which provides the "best forecast" of y_i given the information X_i. Equatio (1) defines the forecast error, ϵ_i.

[A] *The Logit Model*

A typical approach in econometrics is to parameterize the conditiona expectation in (1) as $F(X_i'\alpha)$, where F is some distribution function and α is a vector o parameters, e.g., Amemiya (1985). If F is taken to be the standard normal distributio function, we have the probit model. Here F is taken to be the logistic function,

$$F(X'\alpha) = 1/[1 + \exp(-X'\alpha)], \tag{2}$$

so we have specified the logit model. The parameters α can be estimated by maximizing the following log–likelihood function:

$$\log L = \Sigma_{i=1}^{n} y_i \log F(X_i'\alpha) + (1 - y_i)\log[1 - F(X_i'\alpha)]. \qquad (3)$$

The predicted probability of owning an appliance is then given by $F(X_i'\hat{\alpha})$, where $\hat{\alpha}$ is an estimate of α. It should be emphasized that we do not assume that this is correctly specified. The logistic function (2) is at most an approximation to $E[y_i | X_i]$.

In this paper we use the simple logit model to estimate the conditional mean. It is well known that the logistic distribution is close to the cumulative normal distribution, except at the extreme tails. Therefore, the logit and probit model provide very similar results. In fact, the parameter estimates are in theory comparable when the logit estimates are multiplied by 0.625 (Amemiya (1981)). We do not consider the multinomial logit model here because we are only interested in classifying ownership and nonownership of individual appliances, not ownership of entire portfolios of appliances (cf. Hausman (1979)). Also, we do not adopt the nested logit model because the appliances under analysis are not all related (cf., Dubin (1985, Chap. 3)).

[B] *The Neural Network Model*

The possibility of misspecification motivates us to find an alternative model that can perhaps better approximate the conditional mean. An interesting class of approximating functions is the class of multi– layer feedforward neural network models. This class of functions is capable of approximating broad classes of functions to any desired degree of accuracy (Hornik, Stinchcombe, and White (1989)). It seems reasonable to expect that neural network models can do well in this ownership classification problem.

Let the network "output" o_i be given by the following equations, which define a "single hidden layer feedforward network":

$$o_i = G(\beta_0 + A_i'\beta) = G(\beta_0 + \Sigma_{j=1}^{q} a_{ij}\beta_j)$$

$$a_{ij} = \Psi(X_i'\gamma_j) = \Psi(\gamma_{j0} + \Sigma_{k=1}^{p} x_{ik}\gamma_{jk}),\ j=1,\cdots,q, \tag{4}$$

where $A_i = (a_{i1},\cdots,a_{iq})$ is vector of "hidden unit activations," X_i is a vector of inputs (explanatory variables) inclduing a constant term, $\beta = (\beta_1,\cdots,\beta_q)'$ and $\gamma_j = (\gamma_{j0},\cdots,\gamma_{jp})'$, j=1,$\cdots$,q, are parameters ("network connection weights"), and G, Ψ are some known functions. That is, inputs (demographic variables) first activate each hidden unit in the intermediate layer through the function Ψ, and activations of hidden units in turn affect outputs through the function G.

In this paper we choose Ψ as the logistic function and G as the identity function. This choice is convenient and suffices for the desired approximation property. Note that the logistic function is a continuous version of the threshold function. Hence the function Ψ in the network plays the role of classifier which characterizes nonlinear features of the function to be approximated. The more hidden units are available in the network, the better approximation the network can produce. From (4) we obtain

$$o_i = \beta_0 + \Sigma_{j=1}^{q}\Psi(\gamma_{j0} + \Sigma_{k=1}^{p} x_{ik}\gamma_{jk})\beta_j = f(X_i,\theta), \tag{5}$$

where $\theta \equiv (\beta_0,\cdots,\beta_q,\gamma_1',\cdots,\gamma_q')'$. In our application, we fix the number of hidden units (q = 4) so that the function f in (5) can only approximate unknown functions to a fixed degree of accuracy. Nevertheless, f appears to be a reasonable approximating function to the conditional mean function, and the network outputs o_i should match $E[y_i | X_i]$ fairly closely.

The parameters θ in the network (5) are estimated by the method of nonlinear least squares (NLS). The predicted probability of owning an appliance is then given by $f(X_i,\hat{\theta})$. The most commonly used estimation method associated with feedforward neural network models is the "back–propagation" estimator (Rumelhart, Hinton, and Williams (1986)). This method is a recursive estimation scheme implementing a gradient search over the parameter space. The back–propagation method, like the gradient method in numerical optimization, may converge very slowly (e.g., White (1988)), but it is appealing when on–line data are available. However, we do not use the back–propagation estimator because the Miracle data sets are not on–line data. Instead, we use the method of NLS. The NLS estimates are consistent and asymptotically normally distributed under general conditions, even in misspecified models, see e.g., White (1990). They are also asymptotically efficient relative to back–propagation estimates (White (1989)).

[C] *Regression Tree Analysis*

The third methodology for classifying owners and nonowners of appliances is regression tree analysis (Breiman, Friedman, Olshen, and Stone (1984)). This technique performs a sequence of binary splits according to household characteristics (demographic variables) and results in a "tree" structure for classifying appliance ownerships. The regression tree analysis differs from the other two models discussed in the preceding subsections in that it is a nonparametric technique. Unlike other nonparametric procedures such as the kernel estimation, the regression tree analysis provides information regarding the structure of the data, as in the standard regression analysis.

In the beginning of the tree creation process, the whole data set belongs to a root node. The regression tree method iteratively performs binary splits according to some household characteristics x_{im} (an element of X_i) so that each X_i can be assigned to either one of the descendent nodes. Let there be N observations, and define $\mathcal{X} \subseteq \mathbb{R}^k$ as the measurement space such that $X_i \in \mathcal{X}$ for all i. Creating a tree is equivalent to partitioning the space $\mathcal{X}$ into different "rectangles". In what follows, T denotes a tree, t denotes a node in the tree, and $\tilde{T}$ denotes the set of terminal nodes in the tree. Hence t is a subset of $\mathcal{X}$, and $\tilde{T}$ forms a partition of $\mathcal{X}$. Define the average of y_i within node t as

$$\bar{y}(t) = \frac{1}{N(t)} \Sigma_{X_i \epsilon t}\, y_i, \tag{6}$$

where N(t) is the number of observations in node t, and define the error measure at node t as

$$R(t) = \frac{1}{N} \Sigma_{X_i \epsilon t}(y_i - \bar{y}(t))^2. \tag{7}$$

From each node t, a candidate split s is such that, for some cut–off value c, we have left and right descendent nodes:

$$t_L = \{X_i\text{: the mth coordinate } x_{im} \le c\},$$

and

$$t_R = \{X_i\text{: the mth coordinate } x_{im} > c\}.$$

The best split s^* is defined to be the split such that

$$\Delta R(s^*,t) = \max_{s\epsilon S} \Delta R(s,t)\ ,$$

where S is the set of all candidate splits, and

$$\Delta R(s,t) = R(t) - [R(t_L) + R(t_R)].$$

That is, the best split maximizes the decrease of the error among all candidate splits. Therefore, a node can be successively split into descendent nodes, and a "tree" type structure can be constructed.

It can be shown that $R(t) \ge R(t_L) + R(t_R)$ for any split. Define the error measure of the tree T as the sum of error measures of all terminal nodes $\tilde{T}$, i.e.,

$$R(T) = \Sigma_{t \epsilon \tilde{T}} R(t), \tag{8}$$

where R(t) is given by (7). Clearly, $R(T) \ge R(T')$ if T' is grown from T. Therefore, we tend to do more splitting and grow a very large tree if R(T) is used as a performance criterion. Consequently, we tend to make every terminal node "pure". This is analogous to the problems created by adding ever more explanatory variables to a regression function.

We can overcome this "over–growing" problem by first developing a large tree

T_{max} and then pruning this large tree upward, where T_{max} is determined by setting the minimum number of observations in each terminal node. Consider the following error–complexity measure:

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}|, \tag{9}$$

where $|\tilde{T}|$ is the number of terminal nodes in $\tilde{T}$, and $\alpha > 0$ is the "complexity parameter." In (9), the error measure of a complex tree with many terminal nodes is penalized by the term $\alpha|\tilde{T}|$. The magnitude of penalty depends on the value of α. It can be shown that there is a decreasing sequence of subtrees of T_{max} ($T_{max} \supseteq T_1 \supseteq T_2 \cdots \supseteq$ root node) and a corresponding increasing sequence of α values ($0 = \alpha_1 < \alpha_2 < \cdots$) such that T_j is the smallest subtree of T_{max} minimizing $R_\alpha(T)$, where $\alpha_j \leq \alpha \leq \alpha_{j+1}$. After the sequence $\{T_j\}$ is obtained, we can use cross–validated estimates $R^{cv}(T_j)$ for $R(T_j)$ and choose the optimal subtree T_* by the "1 SE rule". That is, we choose the smallest subtree T_* such that

$$R^{cv}(T_*) \leq [\min_j R^{cv}(T_j)] + SE\ ,$$

where SE is some standard error estimate. The intuition of the 1 SE rule can be found in Breiman, Friedman, Olshen, and Stone (1984, pp.78–80). The details of growing and pruning a tree can also be found in the same book. All the procedures described above are implemented by the program CART (Classification And Regression Tree).

Once an optimal tree is constructed, each terminal node is assigned as owner or non–owner by the plurality rule. That is, a terminal node is an "owner" node if there are more owners than non–owners falling into this node. A new observation X_n now can be easily classified into owner or nonowner by running X_n through the tree structure and checking which terminal node the new observation ends up with. Alternatively, we can assign an estimate of the probability of ownership for a household belonging to a given terminal node as equal to the proportion of owners belonging to that terminal node, i.e.,

$\bar{y}(t)$, where $t \in \tilde{T}$. We note that the appliance ownership problem is a binary choice problem. Hence the regression tree is virtually the same as the two–class classification tree discussed in Breiman et. al. (1984).

3. The Data and Computer Programs

The data used in this paper comprise part of the Miracle 4, 5, and 6 datasets collected by SDG&E. The Miracle 4 survey was conducted in 1979 and yielded 12,380 usable observations; the Miracle 5 survey was sent out in 1981 and resulted in 8022 usable observations; the Miracle 6 survey was conducted in 1983 and resulted in 7600 usable observations. We specifically utilize information about household appliance ownership and consumer demographics. Observations are usable if ownership and certain (but not all) values of the explanatory variables are not missing.

In this study we focus on 7 gas appliances and 15 electric appliances. The gas appliances under analysis include: (1) range; (2) dryer; (3) water heater; (4) main heating system; (5) air conditioner; (6) fireplace; (7) B.B.Q. Data for the last two gas appliances are not available in the Miracle 4 data set. The electric appliances consist of: (1) black and white TV; (2) color TV; (3) dishwasher; (4) microwave oven; (5) range; (6) dryer; (7) washer; (8) refrigerator; (9) water heater; (10) main heating system; (11) air conditioner; (12) attic fan; (13) air cleaner; (14) electric blanket; (15) water bed. Data for the last four electric appliances are not available in the Miracle 4 data set.

There are eight demographic variables used to characterize appliance ownership: (1) home ownership (Nhomeown); (2) age of dwelling unit (Nyrbuilt); (3) number of bedrooms (Nbedroom); (4) square footage of residence (zsqfoot); (5) number of persons in household (znuminhh); (6) educational attainment in years of head of household (Neducate); (7) family income (zincome); (8) type of dwelling unit (Nresid).

For each data set, the appliance ownership dummy variables are transformed from raw survey data into binary variables with values 1 and 0, indicating owner and non–owner, respectively. Some demographic variables, e.g., square footage and household income, are transformed into the midpoints of the ranges given by the survey questions. For example, a household income is assigned \$22,500 if the survey response indicates the income is within the range \$20,000–\$24,999. Some observations are dropped because of missing values or inappropriate responses. However, missing information for certain variables is assigned the average value of the valid observations. A detailed description of the data transformation can be found in Granger, Kuan, Mattson, and White (1989).

The logit model is estimated using "Statistical Software Tools" (SST) version 1.8 by J. A. Dubin and R. D. Rivers. The regression tree is created using the CART program version 1.1 by California Statistical Software, Inc. The neural network models are estimated by the method of NLS. The γ connection weights are initialized randomly, and the β, γ weights are then adjusted iteratively to minimize the average of squared errors.

4. Overview of Results

In this section we discuss and compare the empirical results obtained from the logit analysis, the neural network analysis, and the regression tree analysis.

The sample averages of appliance–ownership dummy variables are summarized in Tables 1 and 2. It is easily seen that these values change a lot from Miracle 4 to Miracle 5 but remain relatively stable from Miracle 5 to Miracle 6. This may be due to the fact that the questionnaire used for the Miracle 4 survey is quite different from the other two surveys, and that the Miracle 5 and 6 surveys are subject to the survey requirements imposed by California Energy Commission. We observe an exception in that the sample average of the black and white TV ownership variable drops from .39 (in Miracle 4) to .288

(in Miracle 5) and then rises to .874 (in Miracle 6). We notice that only 2000 observations for black and white TV are valid in Miracle 6, in contrast with 6700–6900 valid observations for other appliances. It is likely that most of the non–owners are excluded because of missing values. Thus the results for black and white TV are likely to be unreliable.

Models for each appliance in each data set are estimated separately in this study. We consider both in–sample and out–of–sample predictions. The out–of–sample predictions are obtained by substituting the Miracle 5 and 6 data into the models estimated with the Miracle 6 and 5 data, respectively. We do not use the Miracle 5 or 6 data to evaluate the model estimated with the Miracle 4 data because of the incompatibility of the questionnaires in these surveys.

An example of the estimation results for each of the models is given in Tables 3A, B, and C. The particular results given are for Miracle 5, electric main heating system. Similar results for each sample and each appliance are available from the authors on request. Because our interest centers on comparing the different methods, we do not provide a detailed analysis of the results of individual estimated models (there are a total of 180 estimated models), but instead turn our attention to comparisons of model performance.

The criterion we use to compare the performance of models is the average of log–likelihood values. For the logit and neural network models the average is calculated by:

$$N^{-1}\Sigma_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i), \qquad (10)$$

where $\hat{y}_i$ is the predicted value and N is the number of valid observations. For the logit model, $\hat{y}_i = F(X_i'\hat{\alpha})$ is calculated from (2), and $\hat{\alpha}$ is the vector that maximizes (3). For the neural network model,

$$\hat{y}_i = f(X_i,\hat{\theta}), \text{ if } .001 \le f(X_i,\hat{\theta}) \le .999 \qquad (11)$$
$$= .999, \quad \text{if } f(X_i,\hat{\theta}) > .999$$
$$= .001, \quad \text{if } f(X_i,\hat{\theta}) < .001,$$

where $f(X_i,\hat{\theta})$ is calculated from (5), and $\hat{\theta}$ is the NLS estimator.

In the regression tree analysis, each observation is assigned to a terminal node, and a probability of ownership is assigned as $\bar{y}(t)$. Recall that $\bar{y}(t)$ denotes the sample average of y_i within node t (as in (6)), N(t) denotes the number of observations in node t, and $|\tilde{T}|$ is the number of terminal nodes. Putting $\hat{y}_i = \bar{y}(t)$ for $X_i \in t$, the average of log–likelihood values is calculated by

$$N^{-1}\Sigma_{i=1}^{N} y_i\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i) =$$

$$N^{-1}\Sigma_{t=1}^{|\hat{T}|} \bar{y}(t)N(t)\log(\bar{y}(t)) + (1-\bar{y}(t))N(t)\log(1-\bar{y}(t)). \qquad (12)$$

We note that not all demographic variables are used to create a regression tree. A demographic variable is used for splitting only when such a split can improve upon the error measure. Hence the regression tree for each appliance is different. In some extreme cases, there is no tree created because of the very high (low) sample averages and the "1 SE rule". If there is no tree created, $|\hat{T}| = 1$ and $\bar{y}(t)$ is the sample average of all y_i. The in–sample averages of log–likelihood values are given in Tables 5 and 6, and the out–of–sample averages are listed in Tables 8 and 9. Tables 4A and B summarize this information and give the number of best performances of each model for every data set.

The average of log–likelihood values is an appropriate criterion for evaluating the performance of different models, as the summand in Equation (10) measures the "entropy" of the estimated distribution relative to the true distribution (see e.g., Theil (1971, pp. 636–640)). If $\hat{y}_i$ is close to zero (one) and $y_i = 0$ (1), the prediction is accurate and the summand in (10) is close to zero. On the other hand, if $\hat{y}_i$ is close to zero (one) but $y_i = 1$ (0), the prediction is very poor and the summand in (10) is very negative. Thus, the sum in (10) measures the total "surprise" resulting from the contradiction between the predicted probabilities and the true outcomes. A model that performs better should yield less "surprise", compared to the other models. Equation (12) is interpreted in a similar fashion. The average values allow for the comparison across appliances and surveys, since the number of valid observations N differs for each appliance.

The truncation for the predicted probabilities in (11) is needed to ensure proper calculation of the log–likelihood. When the predicted probability is outside the range [.001, .999], the resulting likelihood will be underestimated if the true outcome is opposite. However, very few observations fall in this category, and this number is less than 20 for most of the in–sample and out–of–sample forecasts. Table 10 gives some examples of the worst cases for out–of–sample forecasts, in which the number of misclassification is shown in the off–diagonal entries.

From the summary statistics in Tables 4A and 4B we can see that the neural network model outperforms the other two models in–sample; out–of–sample, the regression tree performs better for gas appliances, and all three models have similar performance for electric appliances. A detailed comparison can be made by using the information in Tables 5, 6, 8 and 9.

Tables 5 and 6 contain, respectively, the in–sample averages of log–likelihoods for gas and electric appliances in each data set. We observe that, when the neural network model is dominated by the other models, the difference between the average values of the network and the best model is typically small. For gas appliances, the largest difference is .0016, and most of the differences are below .001. For electric appliances, the largest difference is .0019, and most of the differences are around .0015. On the other hand, when the neural network model dominates the other models, its average value usually differs from that of the second best model by a larger amount. Some of the differences are greater than .01. This shows that the neural network model outperforms the other models significantly in–sample. It can also be seen that the logit model performs better than the regression tree model.

In order to determine whether the in–sample differences reported in Tables 5 and 6 are statistically significant, we compute a version of Vuong's (1989) statistic for model selection of strictly non–nested models. The version of Vuong's statistic computed here can be expressed as

$$V_N = 1/N \,\Sigma_{i=1}^{N}(g_i - h_i)^2 - [1/N \,\Sigma_{i=1}^{N}(g_i - h_i)]^2$$

where

$$g_i = y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)$$

is individual log–likelihood obtained from the network model with $\hat{y}_i$ calculated from (11) and

$$h_i = y_i \log(F(X_i'\hat{\alpha})) + (1 - y_i)\log(1 - F(X_i'\hat{\alpha}))$$

if we compare the network and the logit model or

$$h_i = y_i \log(\bar{y}(t)) + (1 - y_i)\log(1 - \bar{y}(t)),\ X_i \in t,$$

if we compare the network and the regression tree model.

Under the null hypothesis that the two models compared (e.g., the neural network model and the logit model) have equal expected log–likelihood, Theorem 5.1 of Vuong (1989) establishes that this statistic is asymptotically distributed as standard normal. The values for these statistics, comparing the neural network model to the logit and CART models respectively for Miracle 5 and 6 are given in Table 7A for gas appliances and Table 7B for electric appliances. For example, the Vuong statistic is 4.497 for the neural network vs. the logit model of gas range ownership in the Miracle 5 data. This has a one–sided p–value (probability of wrongly rejecting the null hypothesis against the alternative of superior performance by the network model) of practically 0. For the CART model, the Vuong statistic is .552, implying a one–sided p–value of .709.

Looking over the results of Tables 7A and 7B, we see that of the cases in which the network model exhibits superior performance, this superiority is statistically significant at the standard 5% level except for gas range, dishwasher, microwave, electric range, and washer in Miracle 5 and except for washer, main heating and air cleaner in Miracle 6.

In out–of–sample predictions for gas appliances, the regression tree model turns out to perform best. The problem with the tree model is that its performance is rather bad when no tree is created, as for gas air conditioner and gas B.B.Q. We also observe from Table 8A that the neural network model never outperforms the other models for gas appliances, but it is the second best model for 5 out of 7 gas appliances. It is also interesting to see that for some appliances (gas dryer, water heater and main heating), the out–of–sample average values of the network are better than the in–sample averages of the logit model. In Table 8B the neural network model is the best model for 2 out of 7 gas appliances, and for the other appliances it is the worst model. In both cases, the logit model is always the best for gas B.B.Q., and the tree model is always the best for gas range, water heater and main heating.

Tables 9A and 9B contain the out–of–sample averages of log–likelihood values for electric appliances. In Table 9A, the regression tree model is the best (second best) for 5 (6) out of 15 appliances; and the neural network model is the best (second best) for 5 (4) appliances. In Table 9B, the regression tree model is the best (second best) for 7 (4) appliances; and the network is the best (second best) for 4 (6) appliances. In both cases, the regression tree model always performs well for electric range, washing machine, and main heating system; the network is always the best for electric dryer and electric blanket;

and the logit model always performs well for microwave, refrigerator, and air cleaner. We also note that the CART program does not create a tree for 6 out of 15 appliances in these two tables. As for the results for gas appliances, the performance of the tree model is usually poor when no tree is created. The exceptions are water heater in Table 9A and attic fan in Table 9B, for which guesses yield better log–likelihood values.

Intuitively, the out–of–sample likelihood values should be worse than the in–sample values. This is true for the logit model. We observe the following exceptions for the regression tree model: gas dryer and gas water heater in Table 8A and electric air conditioner and water bed in Table 9B. There is also one exception for the neural network model: refrigerator in Table 9B. Local rather than global optimization in sample explains these results.

Our results indicate that the tree model can do well in out–of–sample contexts. However, if there is no tree created for an appliance and the sample averages of that appliance are quite different in two data sets, the in–sample and out–of–sample likelihood values differ significantly. For example, the difference of likelihood values for electric water heater is 0.062 in Table 9A and 0.123 in Table 9B. Another interesting example is that of color TV. The out–of–sample likelihood value resulting from the tree created in Miracle 5 is very close to the in–sample value (see Table 9A). But there is no tree created for color TV in Miracle 6, hence the out–of– sample likelihood value differs from the in–sample value by 0.203 in Table 9B. These facts also suggest that the regression tree may not be very useful for out–of–sample forecasting when no tree can be created.

5. Summary and Concluding Remarks

In this empirical study we find that the prediction ability of two novel methods using neural network and regression tree models is reasonably good for this classification problem. Although the neural network model does not uniformly dominate the logit and the regression tree models, it does outperform these models in in–sample prediction of ownerships of many appliances. For out–of–sample prediction, the regression tree model is most successful for gas appliances, but its ability is weakened when the CART program fails to create a tree. The network and logit model also perform reasonably well out of sample. The price paid for the increased performance of the regression tree and neural network models is that they are computationally more intensive to estimate than the logit model.

Although the results reported here are informative, they cannot be the last word. Instead, they suggest the usefulness of further study of the relative performance of the network and CART models, given that the network models have better in–sample performance and the CART models have better out–of–sample performance. An obvious reason for the better out–of–sample performance for CART is its use of cross– validation to determine the optimal tree structure. Similar use of cross–validation to determine the optimal number of hidden units (currently fixed at four in this study) may be expected to lead to further improvements in out–of–sample performance for the network models. Because of the huge computational effect required, convenient cross–validation methods for nonlinear network models are not presently available. Development of such methods is now in progress.

Another source of possible improvement in both in and out of sample network performance is use of a "squashing function" at the output unit, achieved by replacing the present choice of G (the identity function) in equation (4) with a function such as the logistic (already used for Ψ). This forces the network toward making more definite classifications and eliminates problems with outputs greater than one or less than zero. Associated with this replacement is use of minimum entropy quasi–maximum likelihood estimation in place of current NLS techniques. This too should lead to further improvements in both in– and out–of–sample network performance.

Performance of the regression tree model may also be improved by experimenting with the CART program options. For example, we may decrease the minimum size below which nodes will not be split, we may use linear combination of variables, and we may use the "zero SE rule" instead of the "one SE rule" to select the tree. We leave investigation of these possibilities to further research.

Table 1 Sample Proportions of Gas-Appliance Ownership.

Appliance	Miracle 4	Miracle 5	Miracle 6
Range	.487	.484	.480
Dryer	.343	.291	.307
Water Heater	.793	.656	.655
Main Heating	.790	.684	.686
Air Conditioner	.018	.016	.019
Fireplace	N/A	.199	.217
B.B.Q.	N/A	.051	.046

N/A: Not available.

Table 2 Sample Proportions of Electric-Appliance Ownership.

Appliance	Miracle 4	Miracle 5	Miracle 6
B/W TV	.390	.288	.874
Color TV	.896	.884	.990
Dish Washer	.600	.584	.572
Microwave	.308	.365	.441
Range	.506	.496	.488
Dryer	.361	.305	.295
Washing Machine	.785	.692	.703
Refrigerator	.996	.976	.977
Water Heater	.111	.091	.010
Main Heating	.172	.152	.159
Air Conditioner	.097	.216	.213
Attic Fan	N/A	.049	.048
Air Cleaner	N/A	.029	.036
Elec. Blanket	N/A	.410	.389
Water Bed	N/A	.163	.144

N/A: Not available.

Table 3 Estimation Results for Electric Main Heating System in Miracle 5.

A. The Logit Model:

Variables	Coefficient	Standard Error
Constant	-.651913	.297783
Nhomeown	-.214032	.079322
Nyrbuilt	-.067313	.004298
Nbedroom	-.353786	.052370
Neducate	.064085	.020598
zsqfoot	.000193	.000077
znuminhh	-.010022	.031496
zincome	.000012	.000003
Nresid	-1.061678	.090408

Initial Likelihood: 5258.2. Likelihood at convergence: -2762.7.

B. The Neural Network Model:

Input Variables	Gamma Weights Connecting Input Units to Hidden Units #1	#2	#3	#4
Constant	.902074	-.555334	1.654142	2.276281
Nhomeown	-4.181901	.444344	1.173967	1.018036
Nyrbuilt	-7.418556	-1.149316	-1.050310	.941213
Nbedroom	.589866	-1.811435	1.790252	1.737879
Neducate	-3.996109	.890788	-1.133193	.632603
zsqfoot	-1.617610	.138512	.385821	-.835286
znuminhh	2.524449	-.091295	1.992484	1.138858
zincome	2.604609	.087547	1.547868	1.353484
Nresid	-5.683916	-.071938	2.302176	.930138

Beta Weights Connecting Hidden Units to Output Units

Bias	#1	#2	#3	#4
-.913163	.410989	.832218	.043323	.771775

C. The Regression Tree:

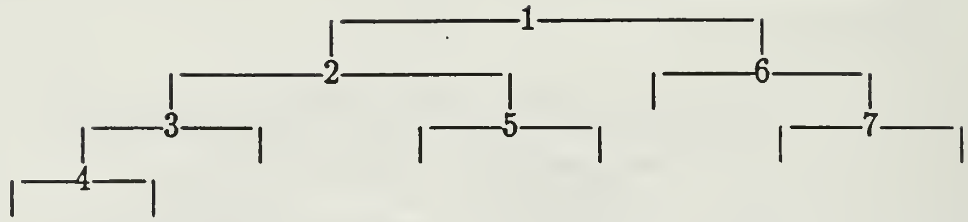
Options Used:

1. Construction Rule: Least Squares
2. Estimation Method: 10-fold cross validation
3. Tree Selection Rule: 1 SE Rule
4. Linear Combinations: No
5. Initial value of the complexity parameter = 0.0
6. Size requirement for subsampling = 1000
7. Minimum size below which node will not be split = 200
8. Maximum number of surrogates used for missing values = 7
9. Maximum number of nodes in largest tree grown = 150
   (Actual number of nodes in largest tree grown = 90)
10. Maximum depth of largest tree grown = 250
   (Actual maximum depth of largest tree grown = 16)
11. Maximum size of memory available = 150000
   (Actual size of memory used in run = 110929)

Tree Sequence:

Tree	Terminal Nodes	Cross-Validated Relative Error	Resubstitution Relative Error	Complexity Parameter
53	12	.86 +/- .010	.84	2.87
54	11	.86 +/- .010	.84	2.97
55	10	.86 +/- .010	.84	3.18
56	8	.86 +/- .010	.85	3.71
57	7	.87 +/- .009	.86	4.67
58	6	.87 +/- .009	.86	6.57
59	5	.88 +/- .009	.87	10.5
60	3	.90 +/- .008	.90	12.9
61	2	.94 +/- .005	.93	28.0
62	1	1.00+/- .000	1.00	69.0

Tree Diagram:



Terminal Regions: 1 2 3 4 5 6 7 8

Split Information:

Split #1 on variable Nresid
Split #2 on variable Nyrbuilt
Split #3 on variable Nhomeown
Split #4 on variable zsqfoot
Split #5 on variable Nyrbuilt
Split #6 on variable Nhomeown
Split #7 on variable Nyrbuilt

Terminal Node Information:

Node	Cases	Average*	SD
1	439	.535	.50
2	434	.362	.48
3	527	.245	.43
4	1151	.234	.42
5	413	.029	.17
6	2851	.045	.21
7	975	.176	.38
8	796	.060	.24

* Average = percentage owning appliance in terminal node.

Table 4A Model Performance Comparison for Gas Appliances.

# of best performance	In Sample			Out of Sample	
	M-4 data	M-5 data	M-6 data	M-6 data M-5 model	M-5 data M-6 Model
Logit	1	2	2	3	1
CART	1	0	2	4	4
Network	3	5	3	0	2

M-4 (5,6) stands for Miracle 4 (5,6).

Table 4B Model Performance Comparison for Electric Appliances.

# of best performance	In Sample			Out of Sample	
	M-4 data	M-5 data	M-6 data	M-6 data M-5 model	M-5 data M-6 Model
Logit	2	3	3	5	4
CART	0	1	1	5	7
Network	9	11	11	5	4

M-4 (5,6) stands for Miracle 4 (5,6).

Table 5A In-Sample Averages of Log-Likelihoods: Miracle 4 Gas Appliances.

Appliance	Logit	CART	Network	N
Range	−.6069	-.6015	−.6012*	11841
Dryer	−.5862	−.5851	−.5809*	11709
Water Heater	−.4352	−.4234*	−.4250	11501
Main Heating	−.4612	−.4525	−.4475*	11620
Air Conditioner	−.0853*	−.0901†	−.0857	11663

Note: * Best performance. † No tree created.

Table 5B In-Sample Averages of Log-Likelihoods: Miracle 5 Gas Appliances.

Appliance	Logit	CART	Network	N
Range	−.6272	-.6172	−.6154*	7597
Dryer	−.5387	−.5273	−.5218*	7654
Water Heater	−.5627	−.5438	−.5378*	7569
Main Heating	−.5671	−.5627	−.5520*	7586
Air Conditioner	−.0764*	−.0820†	−.0778	7453
Fireplace	−.4050	−.4129	−.3977*	7396
B.B.Q.	−.1885*	−.2014†	−.1894	7394

Note: * Best performance. † No tree created.

Table 5C In-Sample Averages of Log-Likelihoods: Miracle 6 Gas Appliances.

Appliance	Logit	CART	Network	N
Range	-.6062	-.6004*	-.6010	6910
Dryer	-.5363	-.5354	-.5266*	6814
Water Heater	-.5396	-.5265	-.5140*	6760
Main Heating	-.5442	-.5260*	-.5272	6789
Air Conditioner	-.0852*	-.0941†	-.0861	6750
Fireplace	-.4196	-.4261	-.4083*	6916
B.B.Q.	-.1650*	-.1866†	-.1654	6912

Note: * Best performance. † No tree created.

Table 6A In-Sample Averages of Log-Likelihoods: Miracle 4 Electric Appliances.

Appliance	Logit	CART	Network	N
B/W TV	-.6591	-.6610	-.6567*	11749
Color TV	-.2940	-.3043	-.2906*	11569
Dish Washer	-.4690	-.4654	-.4644*	11684
Microwave	-.5564	-.5578	-.5504*	11731
Range	-.6045	-.5987	-.5984*	11841
Dryer	-.6002	-.5996	-.5875*	11709
Washing Machine	-.2955	-.3054	-.2929*	11635
Refrigerator	-.0237*	-.0261†	-.0239	11868
Water Heater	-.3362	-.3486†	-.3310*	11501
Main Heating	-.3925	-.3882	-.3860*	11620
Air Conditioner	-.2929*	-.3013	-.2939	11663

Note: * Best performance. † No tree created.

Table 6B In-Sample Averages of Log-Likelihoods: Miracle 5 Electric Appliances.

Appliance	Logit	CART	Network	N
B/W TV	-.5921	-.6004†	-.5874*	7710
Color TV	-.3232	-.3401	-.3204*	7721
Dish Washer	-.4989	-.4902	-.4868*	7420
Microwave	-.5756	-.5732	-.5727*	7409
Range	-.6227	-.6122	-.6090*	7597
Dryer	-.5541	-.5570	-.5397*	7654
Washing Machine	-.3534	-.3498	-.3465*	7645
Refrigerator	-.1050*	-.1132†	-.1069	7636
Water Heater	-.2989	-.3048†	-.2923*	7569
Main Heating	-.3642	-.3584*	-.3624	7586
Air Conditioner	-.5135	-.5218†	-.5114*	7453
Attic Fan	-.1896*	-.1956†	-.1911	7395
Air Cleaner	-.1259*	-.1312†	-.1274	7397
Elec. Blanket	-.6515	-.6576	-.6458*	7422
Water Bed	-.4318	-.4337	-.4255*	7387

Note: * Best performance. † No tree created.

Table 6C In-Sample Averages of Log-Likelihoods: Miracle 6 Electric Appliances.

Appliance	Logit	CART	Network	N
B/W TV	−.3727	−.3787†	−.3653*	2057
Color TV	−.0545*	−.0560†	−.0563	6367
Dish Washer	−.4689	−.4685	−.4549*	6919
Microwave	−.6008	−.6025	−.5952*	6924
Range	−.5999	−.5963	−.5878*	6910
Dryer	−.5446	−.5409	−.5262*	6814
Washing Machine	−.3177	−.3212	−.3141*	6826
Refrigerator	−.1030*	−.1095†	−.1041	6685
Water Heater	−.0567	−.0560†*	−.0569	6760
Main Heating	−.3601	−.3572	−.3542*	6789
Air Conditioner	−.5092	−.5112	−.5035*	6750
Attic Fan	−.1814*	−.1926†	−.1830	6911
Air Cleaner	−.1510	−.1521†	−.1508*	6910
Elec. Blanket	−.6359	−.6441	−.6328*	6909
Water Bed	−.3955	−.3993	−.3879*	6907

Note: * Best performance. † No tree created.

Table 7A Vuong's Statistic for Non-nested Models: Network Model vs. Logit and CART, Gas Appliances. (One-sided P-values in parentheses.)

Gas Appliance	Miracle 5 vs. Logit	Miracle 5 vs. CART	Miracle 6 vs. Logit	Miracle 6 vs. CART
Range	4.497	.552	1.405	−.115
	(.000)	(.709)	(.079)	(.548)
Dryer	9.395	3.416	5.171	4.124
	(.000)	(.0003)	(.000)	(.000)
Water Heater	8.867	2.519	8.838	4.399
	(.000)	(.006)	(.000)	(.000)
Main Heating	8.154	4.717	6.854	−.424
	(.000)	(.000)	(.000)	(.663)
Air Conditioner	−1.782	†	−1.850	†
	(.963)		(.968)	
Fireplace	3.124	4.570	4.972	5.664
	(.001)	(.000)	(.000)	(.000)
B.B.Q.	−.763	†	−.297	†
	(.776)		(.618)	

† No tree created.

Table 7B Vuong's Statistic for Non-nested Models: Network Model vs. Logit and CART, Electric Appliances. (One-sided P-values in parentheses.)

Electric Appliance	Miracle 5		Miracle 6	
	vs. Logit	vs. CART	vs. Logit	vs. CART
B/W TV	5.242	†	3.188	†
	(.000)		(.0007)	
Color TV	2.311	9.045	-2.372	†
	(.010)	(.000)	(.991)	
Dish Washer	4.326	.888	5.319	3.207
	(.000)	(.187)	(.000)	(.0007)
Microwave	1.369	.150	3.305	2.526
	(.085)	(.440)	(.0005)	(.006)
Range	4.870	1.026	4.125	2.279
	(.000)	(.152)	(.000)	(.011)
Dryer	6.096	5.694	7.103	5.316
	(.000)	(.000)	(.000)	(.000)
Washer	2.708	1.093	1.420	1.982
	(.003)	(.138)	(.078)	(.024)
Refrigerator	-2.682	†	-1.591	†
	(.996)		(.944)	
Water Heater	3.431	†	-.430	†
	(.0003)		(.666)	
Main Heating	.714	-1.280	2.416	.929
	(.239)	(.900)	(.008)	(.176)
Air Conditioner	3.599	†	4.139	4.407
	(.0002)		(.000)	(.000)
Attic Fan	-2.125	†	-2.069	†
	(.983)		(.981)	
Air Cleaner	-1.572	†	.509	†
	(.942)		(.305)	
Blanket	5.209	5.226	1.949	4.323
	(.000)	(.000)	(.026)	(.000)
Water Bed	4.156	4.181	5.138	5.392
	(.000)	(.000)	(.000)	(.000)

† No tree created.

Table 8A Out-of-Sample Averages of Log-Likelihoods: Miracle 6 Gas Appliances with Miracle 5 Model.

Appliance	Logit	CART	Network	N
Range	-.6157 (-.6062)	-.6092* (-.6004)	-.6179 (-.6010)	6910
Dryer	-.5464 (-.5363)	-.5337*§ (-.5354)	-.5355 (-.5266)	6814
Water Heater	-.5474 (-.5396)	-.5233*§ (-.5265)	-.5324 (-.5140)	6760
Main Heating	-.5481 (-.5442)	-.5347* (-.5260)	-.5414 (-.5272)	6789
Air Conditioner	-.0872* (-.0852)	-.0947† (-.0941)	-.0898 (-.0861)	6750
Fireplace	-.4314* (-.4196)	-.4359 (-.4261)	-.4406 (-.4083)	6916
B.B.Q.	-.1687* (-.1650)	-.1868† (-.1866)	-.1705 (.1654)	6912

Note: * The best performance. † No tree created.
§ Out-of-sample likelihood better than in-sample likelihood.
Numbers in parentheses are in-sample (Miracle 6) averages.

Table 8B Out-of-Sample Averages of Log-Likelihoods: Miracle 5 Gas Appliances with Miracle 6 Model.

Appliance	Logit	CART	Network	N
Range	–.6355 (–.6272)	-.6286* (–.6172)	–.6524 (–.6154)	7597
Dryer	–.5580 (–.5387)	–.5629 (–.5273)	–.5335* (–.5218)	7654
Water Heater	–.5739 (–.5627)	–.5558* (–.5438)	–.5813 (–.5378)	7569
Main Heating	–.5712 (–.5671)	–.5627* (–.5627)	–.5789 (–.5520)	7586
Air Conditioner	–.0806 (–.0764)	–.0823† (–.0820)	–.0789* (–.0778)	7453
Fireplace	–.4258 (–.4050)	–.4244* (–.4129)	–.4265 (–.3977)	7396
B.B.Q.	–.1959* (–.1885)	–.2017† (–.2014)	–.2078 (–.1894)	7394

Note: * Best performance. † No tree created.
Numbers in parentheses are in-sample (Miracle 5) averages.

Table 9A Out-of-Sample Averages of Log-Likelihoods: Miracle 6 Electric Appliances with Miracle 5 Model.

Appliance	Logit	CART	Network	N
B/W TV	-1.1775	-1.1308†	-1.0597*	2057
	(-.3727)	(-.3787)	(-.3653)	
Color TV	-.1238	-.0583*	-.1337	6367
	(-.0545)	(-.0560)	(-.0563)	
Dish Washer	-.4852*	-.4880	-.4862	6919
	(-.4689)	(-.4685)	(-.4549)	
Microwave	-.6061*	-.6096	-.6260	6924
	(-.6008)	(-.6025)	(-.5952)	
Range	-.6124	-.6000*	-.6109	6910
	(-.5999)	(-.5963)	(-.5878)	
Dryer	-.5646	-.5586	-.5404*	6814
	(-.5446)	(-.5409)	(-.5262)	
Washing Machine	-.3367	-.3311*	-.3325	6826
	(-.3177)	(-.3212)	(-.3141)	
Refrigerator	-.1042*	-.1095†	-.1060	6685
	(-.1030)	(-.1095)	(-.1041)	
Water Heater	-.1307	-.1184†*	-.1319	6760
	(-.0567)	(-.0560)	(-.0569)	
Main Heating	-.3641	-.3592*	-.3685	6789
	(-.3601)	(-.3572)	(-.3542)	
Air Conditioner	-.5113	-.5179†	-.5084*	6750
	(-.5092)	(-.5112)	(-.5035)	
Attic Fan	-.1860	-.1926†	-.1848*	6911
	(-.1814)	(-.1926)	(-.1830)	
Air Cleaner	-.1537*	-.1558†	-.1577	6910
	(-.1510)	(-.1521)	(-.1508)	
Elec. Blanket	-.6513	-.6492	-.6418*	6909
	(-.6359)	(-.6441)	(-.6328)	
Water Bed	-.3991*	-.4008	-.4024	6907
	(-.3955)	(-.3993)	(-.3879)	

Note: * Best performance. † No tree created.
Numbers in parentheses are in-sample (Miracle 6) averages.

Table 9B Out-of-Sample Averages of Log-Likelihoods: Miracle 5 Electric Appliances with Miracle 6 Model.

Appliance	Logit	CART	Network	N
B/W TV	−1.5579 (−.5921)	−1.5137†* (−.6004)	−1.5226 (−.5874)	7710
Color TV	−.5195 (−.3232)	−.5431† (−.3401)	−.5099* (−.3204)	7721
Dish Washer	−.5260 (−.4989)	−.4990* (−.4902)	−.5371 (−.4868)	7420
Microwave	−.5787* (−.5756)	−.5893 (−.5732)	−.5923 (−.5727)	7409
Range	−.6360 (−.6227)	−.6281* (−.6122)	−.6395 (−.6090)	7597
Dryer	−.6018 (−.5541)	−.5662 (−.5570)	−.5518* (−.5397)	7654
Washing Machine	−.3917 (−.3534)	−.3566* (−.3498)	−.3816 (−.3465)	7645
Refrigerator	−.1060* (−.1050)	−.1132† (−.1132)	−.1065§ (−.1069)	7636
Water Heater	−.4233 (−.2989)	−.4282† (−.3048)	−.4180* (−.2923)	7569
Main Heating	−.3676 (−.3642)	−.3631* (−.3584)	−.3683 (−.3624)	7586
Air Conditioner	−.5139* (−.5135)	−.5184§ (−.5218)	−.5226 (−.5114)	7453
Attic Fan	−.1972 (−.1896)	−.1956†* (−.1956)	−.1971 (−.1911)	7395
Air Cleaner	−.1282* (−.1259)	−.1331† (−.1312)	−.1298 (−.1274)	7397
Elec. Blanket	−.6734 (−.6515)	−.6705 (−.6576)	−.6609* (−.6458)	7422
Water Bed	−.4342 (−.4318)	−.4290§* (−.4337)	−.4580 (−.4255)	7387

Note: * Best performance. † No tree created.
§ Out-of-sample likelihood better than in-sample likelihood.
Numbers in parentheses are in-sample (Miracle 5) averages.

Table 10 Examples of Out-of-Sample Misclassification.

Dish Washer (Miracle 5 Data with Miracle 6 Model)

	Owner	Non-Owner
$\hat{f}_i$<.001	266	39
$\hat{f}_i$>.999	7	237

Gas Water Heater (Miracle 5 Data with Miracle 6 Model)

	Owner	Non-Owner
$\hat{f}_i$<.001	10	1
$\hat{f}_i$>.999	29	83

Washer (Miracle 6 Data with Miracle 5 Model)

	Owner	Non-Owner
$\hat{f}_i$<.001	9	2
$\hat{f}_i$>.999	17	525

Gas Fireplace (Miracle 6 Data with Miracle 5 Model)

	Owner	Non-Owner
$\hat{f}_i$<.001	595	16
$\hat{f}_i$>.999	0	0

Note: $\hat{f}_i = f(X_i, \hat{\theta})$.

References

Amemiya, Takeshi (1981): "Qualitative Response Models: A Survey," *Journal of Economic Literature*, 19, 1483–1536.

Amemiya, Takeshi (1985): *Advanced Econometrics*, Cambridge: Harvard University Press.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone (1984): *Classification and Regression Trees*, Monterey: Wadsworth.

Dubin, Jeffrey A. (1985): *Consumer Durable Choice and the Demand for Electricity*, New York: North–Holland.

Dubin, Jeffrey A., and Daniel L. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345–362.

Granger, Clive W.J., Chung–Ming Kuan, Matthew Mattson, and Halbert White (1989): "Trends in Unit Energy Consumption: the Performance of End–Use Models," *Energy*, forthcoming.

Hausman, Jerry A. (1979): "Individual Discount Rates and the Purchase and Utilization of Energy–Using Durables," *Bell Journal of Economics*, 10, 33–54.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989): "Multi– Layer Feedforward Networks Are Universal Approximators," *Neural Networks*, 2, forthcoming.

Lapedes, A., and R. Farber (1987a): "Genetic Data Base Analysis with Neural Nets," paper presented to the IEEE conference on Neural Information Processing System—–Natural and Synthetic.

Lapedes, A., and R. Farber (1987b): "Nonlinear Signal Processing Using Neural Networks," paper presented to the IEEE conference on Neural Information Processing System—–Natural and Synthetic.

Parti, M., and C. Parti (1980): "The Total and Appliance–Specific Conditional Demand for Electricity in the Household Sector," *Bell Journal of Economics*, 11, 309–321.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams (1986): "Learning Internal Representations bby Error Propagation," in D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Vol. 1, Cambridge: M.I.T. Press, 318–362.

Theil, Henri (1971): *Principles of Econometrics*, New York: Wiley.

Vuong, Quang H. (1989): "Likelihood Ratio Tests for Model Selection and Non–nested Hypotheses," *Econometrica*, 57, 307–334.

White, Halbert (1988): "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns," in *Proceedings of the IEEE International Conference on Neural Networks, San Diego*, San Diego: SOS Printing, pp II: 451–458.

White, Halbert (1989): "Some Asymptotic Results for Learning in Single Hidden Layer Feedforward Network Models," *Jounal of the American Statistical Association*, forthcoming.

White, Halbert (1990): *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press, forthcoming.